



TITLE:

# Data processing pipeline for serial femtosecond crystallography at SACLA( Abstract\_要旨 )

AUTHOR(S):

Nakane, Takanori

---

CITATION:

Nakane, Takanori. Data processing pipeline for serial femtosecond crystallography at SACLA. 京都大学, 2017, 博士(医学)

ISSUE DATE:

2017-01-23

URL:

<https://doi.org/10.14989/doctor.k20071>

RIGHT:

京都大学	博士（医学）	氏 名	中根 崇智
論文題目	Data processing pipeline for serial femtosecond crystallography at SACLA (SACLA における連続フェムト秒結晶学のためのデータ処理パイプライン)		
(論文内容の要旨)			
<p>X線自由電子レーザー XFEL は、従来の放射光と同等以上の光子をフェムト秒程度の短パルスに圧縮して発生できる新しい X 線光源である。この単パルス特性により、蛋白質の微結晶から放射線損傷の影響を受けない常温構造を取得することができる。励起レーザーや二液混合系と組み合わせることで、蛋白質の動的構造変化を捉える時分割実験も可能である。</p> <p>結晶は一発の XFEL パルスによって破壊されるので、数万～数十万の結晶からの独立した回折像を組み合わせる連続フェムト秒結晶学 SFX によってデータ収集が行われる。全ての X 線パルスが結晶に当たるわけではないため、生データはさらにその数倍の枚数になる。これは従来の結晶学の百から千倍の規模である。そのため、人間によるデータ品質の評価は非現実的であり、自動化されたデータ処理パイプラインが不可欠である。また、XFEL の繰り返し頻度は 30～120Hz であり、一枚の画像は約四百万ピクセルである。このデータを遅延なく処理するためには、複数の計算ノードにまたがった並列計算が必要である。これらの点を踏まえ、日本の XFEL 施設である SACLA での SFX 実験を対象としたデータ処理パイプラインを構築した。欧米で用いられている Cheetah と CrystFEL を元に、SACLA の実験系にあわせた修正と、高速化・自動化のための改良を行った。 SACLA のデータ処理にはオンライン系とオフライン系の二系統がある。前者は検出器データをリアルタイムで取得可能だが、計算機的能力やストレージへの出力帯域に制限がある。後者にはその制限はないが、測定が完全に終わるまでデータを取得できない。そこで、パイプラインも二段階に分けて構築した。オンライン・パイプラインは、画像からの回折点検出のみを行い、ヒット率や検出器飽和度を数秒未満の遅延で表示する。オフライン・パイプラインは、回折像が含まれた画像のみを選別し、メタデータと合わせて HDF5 形式で書き出し、CrystFEL による指数付けと積分を行う。パイプラインの実行制御と結果の可視化を行うプログラムも開発した。処理時間短縮のため、パイプラインにはスレッド並列化とノード間並列化を施した。</p> <p>本システムは、SACLA での SFX 実験の 9 割以上で利用されている。ある 4 日間に渡る実験を例にとると、約 20 種の蛋白質結晶から、30Hz で約 510 万枚、41TB の画像が取得された。パイプラインによる選別と圧縮の結果、最終的な出力サイズは 3.2TB まで削減できた。リアルタイム処理は 16 コア 32 論理スレッドの計算機を用いて行い、データ取得から表示までの遅延は数秒未満であった。CPU 使用率は半分未満であり、将来の 60Hz での運用も現実的であることが示された。オフライン処理は 12 コアの計算ノード 16 台を用いて行い、遅延は 3 から 20 分程度であった。パイプラインからのフィードバックは、データ測定に値する蛋白質の選定や時間配分など実験戦略を立てる上で重要であった。パイプラインの出力を元に分子置換や実験的位相決定を行い、ビームタイム中に初期構造を得ること</p>			

<p>ができた。このように、放射光施設における回折実験に迫る、効率的な実験が可能となった。今後の開発課題としては、処理パラメータ調節の自動化・オフライン処理速度の向上を予定している。</p> <p>参考論文として、本パイプラインを利用した研究例として、SFX では初となる、硫黄原子や塩素イオンを利用した単波長異常分散法による実験的位相決定の報告を附した。</p> <p>（論文審査の結果の要旨）</p> <p>X線自由電子レーザー XFEL からの超短パルス X 線による多数の回折像を組み合わせる連続フェムト秒結晶学 SFX では、放射線損傷のない蛋白質構造やその動的変化を観察できる。30 Hz 以上で取得される膨大な回折像の処理は煩雑で時間を要するため自動化された高速処理システムが求められていたが、分子生物学・構造生物学・工学・情報科学の学際領域であり実現困難だった。</p> <p>本研究では日本の XFEL 施設 SACLA のためのデータ処理パイプラインが実験現場での必要性に基づき設計、構築された。スレッドとノード間並列化で従来の約 20 倍のスループットが実現された。質の悪い回折像を棄却し、画素値を整数化してから圧縮する工夫で、データ量も従来の 1 割未満に削減された。事前検討で決定された画像の棄却に関する閾値は実験条件の変化に対し頑健であった。本成果により、従来のプログラムを人間が操作すると実験後 3 日以上を要した解析の多くが自動化・高速化され、試料の良否を測定開始後 3 分以内に判定して実験戦略に反映し、測定完了後約 30 分で構造決定できるほど効率化された。本システムは SACLA での SFX 実験の 9 割以上で活用されている。例えば参考論文では世界で初めて SFX で非修飾蛋白質の実験的位相決定に成功し、データ品質の高さを実証した。</p> <p>以上の研究は、蛋白質の無損傷構造と動的変化の解明に貢献し、生命現象の理解と創薬への応用に寄与するところが多い。</p> <p>したがって、本論文は博士（医学）の学位論文として価値あるものと認める。なお、本学位授与申請者は、平成 28 年 8 月 3 日実施の論文内容とそれに関連した試問を受け、博士として十分な学識と研究能力を有することを確認され、合格と認められたものである。</p>			
要旨公開可能日：            年            月            日			